



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

REC'D 18 NOV 2003

WIPO

PCT

Bescheinigung

Certificate

Attestation

Die angehefteten Unterla-
gen stimmen mit der
ursprünglich eingereichten
Fassung der auf dem näch-
sten Blatt bezeichneten
europäischen Patentanmel-
dung überein.

The attached documents
are exact copies of the
European patent application
described on the following
page, as originally filed.

Les documents fixés à
cette attestation sont
conformes à la version
initialement déposée de
la demande de brevet
européen spécifiée à la
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02368088.7

**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

BEST AVAILABLE COPY



Anmeldung Nr:
Application no.: 02368088.7
Demande no:

Anmeldetag:
Date of filing: 13.08.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

INTERNATIONAL BUSINESS MACHINES CORPORATION

Armonk, NY 10504
ETATS-UNIS D'AMERIQUE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se référer à la description.)

An adaptive resource management method and system

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)
revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

G06F9/46

Am Anmeldetag benannte Vertragsstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR

AN ADAPTIVE RESOURCE MANAGEMENT METHOD AND SYSTEM

Technical field

The present invention relates to the data processing field, and more specifically to a resource management method and corresponding system.

Background art

Management of different types of resources (such as software components, applications or monitoring activities) is a critical issue in a data processing system with a distributed architecture; this problem is particularly acute for resources that have a high level of complexity or are dispersed across a large number of installations. A typical example of a resource management environment is a software distribution application, which is used to upgrade software products installed on a network of workstations.

The resource management environments known in that art are based on an enforcement model (also known as manager/workers model). In this model, the resource management process is entirely controlled by an authority residing at a central site

of the system. The authority defines the desired target state of every resource, which is directly controlled by a corresponding subject distributed in the system. The authority accesses a central repository storing the (assumed) current state for each pair resource/subject, and determines the management actions required to bring the resource to the target state. The management actions are then enforced remotely from the authority on the subject (which is totally passive).

For example, in the software distribution application cited above the authority defines packages including instructions specifying the actions to be carried out on the workstations for installing or removing selected software products. The package is transmitted to the workstations, and the corresponding instructions are interpreted so as to enforce the desired software configuration.

A drawback of the resource management environments known in the art is the lack of any kind of cooperation between the authority and the subjects. This lack of cooperation may bring about inconsistencies when the subjects change their configuration out of the control of the authority. For example, in the software distribution application the instructions of the package are typically conditioned to a series of hardware parameters of the workstation; in this case, a hardware upgrade of the target workstation may change the result of the evaluation of the conditions defined in the package, thereby making the software products installed on the workstation not consistent with its hardware configuration any longer. This lack of cooperation is unacceptable in high dynamic environments, wherein the configuration of the subjects changes frequently.

Moreover, all the proposed solutions require the authority to maintain information about the location of all the subjects; at the same time, the authority must handle the communication with every subject directly. This drawback is

exacerbated in systems having a very high number of subjects, especially when they can change their spatial location freely.

One further problem may arise when the subjects are not available or are off-line. As a matter of fact, the lack of autonomy on the part of the subject requires that each management action must be completed within a pre-set deadline (with the need of fault tolerance, retry, check-point and restart mechanisms). As a consequence, the complexity of the resource management environment is strongly increased.

Summary of the invention

It is an object of the present invention to provide a resource management method that implements an adaptive model.

Moreover, it is another object of the present invention to have the subjects actively participate in the corresponding process.

It is another object of the present invention to support cooperation between the authority and the subjects.

It is yet another object of the present invention to avoid inconsistencies in the subjects.

Furthermore, it is an object of the present invention to couple the authority and the subjects loosely.

It is another object of the present invention to integrate the authority with the subjects at the information level (instead of at the procedural level).

It is yet another object of the present invention to support subjects that are not available or off-line.

The accomplishment of these and other related objects is achieved, in a data processing structure with a distributed architecture including a plurality of subject entities and at last one authority entity, by a resource management method for self-configuring the subject entities, each subject entity belonging to at least one of a plurality of categories and controlling an instance of at least one resource, the at least one authority entity defining a target state of the resources,

wherein the method includes the steps of: publishing, under the control of the at least one authority entity, a plurality of rules each one including an indication of the target state of a resource for a category of the subject entities, each subject entity retrieving the rules corresponding to the at least one category of the subject entity, and applying each retrieved rule to configure the subject entity according to the target state indicated in the rule.

The present invention also provides a computer program application for performing the method, and a program product storing the application.

Moreover, the invention encompasses a computer program running on the authority entity and a computer program running on the subject entity.

The present invention also provides a resource management system implementing the method, a computer of the authority entity and a computer of the subject entity.

The novel features believed to be characteristic of this invention are set forth in the appended claims. The invention itself, however, as well as these and other related objects and advantages thereof, will be best understood by reference to the following detailed description to be read in conjunction with the accompanying drawings.

Brief description of the drawings

- Figure 1a** is a schematic block diagram of a data processing system in which the resource management method of the invention is applicable;
- Figure 1b** shows the functional blocks of a generic computer of the system;
- Figure 2** depicts the main software components used for implementing the resource management method;
- Figures 4a-4d** are activity diagrams describing the flow

of different processes executed by a subject of the system;

Figures 5a-5b is a further activity diagram relating to the operation of a helper of the system;

Figure 6a shows a package diagram grouping the classes used to implement the method;

Figure 6b depicts a corresponding class diagram;

Figure 7 is a sequence diagram exemplifying a resource management operation.

Detailed description of the preferred embodiment

With reference in particular to Figure 1a, a data processing system 100 with a distributed architecture (typically INTERNET-based) is shown. The system 100 implements an environment for managing several kinds of resources, such as products in a software distribution application, activities in a monitoring application, and the like.

A computer 105 operates as an authority, which is responsible to define a desired configuration of the resource management environment. Multiple computers 110a, 110d operate as heterogeneous subjects, which directly control (an instance of) one or more resources to be managed. Each subject 110a, 110d consists of a logic and/or physic entity (such as a laptop, a Personal Digital Assistant (PDA), a SIM card of a mobile telephone, every user of a workstation, and the like).

The authority 105 and the subjects 110a, 110d are coupled through a shared repository 115. The repository 115 stores a set of rules; each rule establishes the desired target state of a resource for a corresponding category of the subjects, defined according to their logic and/or physic characteristics. The rule repository 115 is implemented as a distributed shared memory (DSM), which provides a communication model giving the illusion of a shared memory on top of a message passing system. Preferably, the rule repository 115 consists of a tuple space (which is distributed

on one or more computers); the tuple space includes a collection of fundamental data structures called tuples (each one representing a rule); the tuples are retrieved associatively (that is, according to their content rather than their address). The tuple space implements a cooperation model adhering to the peer-to-peer paradigm, wherein no role separation exists among the entities engaging the tuple space.

Some of the subjects (denoted with 110a) are completely autonomous, and access the rule repository 115 directly. Different subjects (denoted with 110d) depend on one or more subjects 120, which operate as helpers for the dependent subjects 110d; each helper 120 bridges between the rule repository 115 and the associated dependent subjects 110d (in order to retrieve the rules for the associated subjects 110d and then enforce their application). The helpers 120 and the dependent subjects 110d are coupled through the rule repository 115. Particularly, every subject 110d that is not completely autonomous inserts a rule identifying itself into the repository 115; a helper 120 retrieving this rule from the repository 115 is then automatically associated with the corresponding subject 110d.

A computer 125 operates as an administrator, which is responsible to define the membership of the subjects 100a, 110d dynamically. The administrator 125 and the subjects 110a, 110d are coupled through a further shared repository 130 (preferably implemented as a tuple space); the repository 130 stores membership data for the subjects 110a, 110d.

As shown in Figure 1b, a generic computer of the system (authority, subject, helper and administrator) is formed by several units that are connected in parallel to a communication bus 150. In detail, a microprocessor (μ P) 155 controls operation of the computer, a DRAM 160 is directly used as a working memory by the microprocessor 155, and a Read Only Memory (ROM) 165 stores basic code for a bootstrap of the computer. Several peripheral units are further connected to the bus 150 (by means of respective interfaces). Particularly,

a mass memory consists of a magnetic hard-disk 170 and a driver 175 for reading CD-ROMs 180. Moreover, the computer includes input devices 185 (for example, a keyboard and a mouse), and output devices 190 (for example, a monitor and a printer). A network Interface Card (NIC) 195 is used to connect the computer in the system.

Similar considerations apply if the resource management system has a different distributed architecture (for example, based on a LAN), if the computers are connected to each other in a different way, if two or more authorities and/or administrators are provided, if the authority and the administrator consist of a single entity, if different subjects are envisaged, if the shared repositories are implemented with a different technology (even of a non-associative type, such as with relational databases), if the dependent subjects and the helpers are associated in another way, if the computers have a different structure or include other units, and the like.

Considering now Figure 2, the main software components used in the resource management system described above are illustrated. The information (programs and data) is typically stored on the respective hard-disks and loaded (at least partially) into the working memories when the programs are running, together with an operating system and other application programs (not shown in the figure). The programs are initially installed onto the hard-disks from CD-ROMs.

Particularly, the authority includes an editor 203 that is used to define new rules 206. Each new rule 206 is supplied to a module 209, which validates its correctness; for this purpose, the module 209 accesses a log 212 storing all the rules currently in force. When the new rule is in contrast with other rules, the editor 203 is notified accordingly. On the contrary, the new rule is supplied to a publisher 215; the new rule is then stored into the log 212 and provided to the corresponding repository.

The rule repository has an engine 218, which manages insertion of new rules into or deletion of pre-existing rules from its shared memory space 221; the engine 218 further controls extraction of a copy of the rules from the shared memory space 221, in response to corresponding requests from the subjects and the helpers. In addition, the subjects (either directly or through the associated helper) may register themselves with the engine 218 (in order to receive the respective rules as soon as they are available); information about pending requests still to be satisfied for these subjects is stored in a catalog 224.

Each autonomous subject includes a membership controller 227a, which assigns the subject to the category defined by its physic and/or logic characteristics. For this purpose, the membership controller 227a cooperates with one or more plug-in components.

For example, a first component 230a assigns the subject to a category defined according to its physic properties. For example, the physic category is specified by hardware characteristics (such as a hard-disk size, a CPU model, or a working memory capacity) and/or software characteristics (such as installed applications, files or folders). For this purpose, the component 230a relies on a hardware inventory scanner, a software inventory scanner and an application scanner.

A different component 233a statically assigns the subject to a category defined according to its logic properties. For example, the static logic category specifies different groups of users (such as secretaries, managers, developers, and system engineers) or different types of entities (such as desktops, laptops, PDAs or mobile telephones); the static logic category of the subject is derived from an identification code that is input during a log-in procedure or is hardwired.

Another component 236a dynamically assigns the subject to a category defined according to further logic properties. For

example, the dynamic logic category specifies different functional units, departments, network domains, and the like. The dynamic logic categories are set by a configuration module 239 of the administrator. The configuration module 239 controls a database 242, which stores the dynamic logic category associated with each subject of the system (denoted by a corresponding identifier, such as a Globally Unique Identifier or GUID). The configuration module 239 publishes corresponding membership data into the respective repository. The membership repository has an engine 245, which manages a shared memory space 248 (storing the membership data); the component 236a directly interfaces with the engine 245, in order to retrieve the dynamic logic category associated with its identifier.

The membership controller 227a supplies the category associated with the subject (physic component, static logic component and dynamic logic component) to a compliance engine 251a. The compliance engine 251a interfaces with the engine 218 (of the rule repository) to retrieve the rules associated with its category. The retrieved rules are stored into a log 254 and then applied to the subject.

For this purpose, the compliance engine 251a drives one or more applications 255a managing corresponding resources. Each application 255a controls a state catalog 257a; the catalog 257a includes a series of records each one specifying the current state of a corresponding resource under management. The application 255a further accesses a transition table 260; for each resource controlled by the subject and for each pair current state/target state, the transition table 260 stores an indication of one or more management actions required to bring the resource from the current state to the target state.

Likewise, each dependent subject includes a membership controller 227d, which cooperates with a physic component 230d, a static logic component 233d, and a dynamic logic component 236d; the subject further stores one or more

applications 255d with their state catalogs 257d (an additional module, not shown in the figure, is used to insert the rule identifying the subject into the shared memory space 221 of the rule repository, in order to associate the subject with a helper). The membership controller 227d and the state catalog 257d supply the category of the subject and the current state of the resources under management to a compliance engine 251h running on the associated helper. The compliance engine 251h accesses an inventory 263; for each subject controlled by the helper, the inventory 263 stores the corresponding rule log and transition table (in addition to information identifying the subject). The compliance engine 251h controls a module 266 running on the dependent subject, which module 266 drives the applications 255d for enforcing the management actions required to bring each resource of the subject from the current state to the target state.

Similar considerations apply if the whole application (consisting of the programs on the different computers) and the corresponding data are structured in a different manner, if other modules or functions are provided, if the category is defined in a different manner, if each component of the category is based on other characteristics of the subject, if the physic and/or logic characteristics of the subjects are detected in another way, if no state catalog is provided (with the current state of each resource that is detected dynamically), if the subject inventory on the helper stores different information, and the like.

Considering now Figure 3, a flow chart describing the logic of the process of creating a new rule by the authority is illustrated. The process implements a method 300 that starts at block 305. Proceeding to block 310, the new rule is defined specifying the target state of a resource for a corresponding category. The method validates the new rule at block 315; particularly, the method verifies that each subject cannot receive rules defining contrasting target states of the same resource. For this purpose, all the pre-existing rules

relating to the same resource as the new rule are extracted from the log; for each extracted rule, if the target state of the new rule is different from the one of the pre-existing rule the method verifies that the associated categories are not overlapping.

Considering now block 320, if the new rule is correct the rule is published into the corresponding repository at block 320; the new rule is then logged at block 325. Conversely, an error condition is entered at block 330. In both cases, the method ends at the final block 335.

For example, let us consider a rule specifying that a resource APPLICATION must be in a state NON_INSTALLED on every subject of a (physic) category HARD-DISK<10Mbytes. A new rule is defined specifying that the same resource APPLICATION must be in a state INSTALLED on every subject of the (static logic) category SECRETARY. The new rule might cause an unstable situation; in fact, if a subject belongs both to the category HARD-DISK<10Mbytes and to the category SECRETARY, the applications of the two rules would result in a loop wherein the resource APPLICATION is continually installed by the new rule and removed by the pre-existing rule. In this case, the new rule should be amended to specify that the target state INSTALLED for the resource APPLICATION only applies to a category defined as SECRETARY AND NOT(HARD-DISK<10Mbytes)...

Similar considerations apply if the pre-existing rules are not logged on the authority (but they are retrieved from the rule repository dynamically), if a different algorithm is used for validating the new rule, or if no validation is carried out before publishing the new rule (with the correctness of the rules that is verified by the subjects directly before their application).

On the other hand, each (autonomous) subject self-adapts to the respective rules published by the authority according to different policies.

In a first embodiment of the proposed resource management method, the subject operates in a pull mode. As shown in the

activity diagram of Figures 4a-4b, a corresponding management process 400a begins at the black start circle 402 in the swim-lane of the subject. Descending into block 403, the subject retrieves the static logic component of its category. The process then enters a waiting loop at block 404. As soon as a pre-set time-out (for example, 1 minute) has expired, the subject determines the other current components of its category. Particularly the subject requests the dynamic logic component to the membership repository at block 405. In response thereto, the engine of the membership repository retrieves the corresponding membership data from its shared memory at block 406. The process continues to block 407, wherein this membership data is returned to the subject. Referring now to block 408 in the swim-lane of the subject, the returned membership data defines the dynamic logic component of the category of the subject. Hardware and software scans are then executed on the subject at block 410 (assuming that the rules for installing the corresponding components have already been applied), in order to detect the corresponding physic component of its category; the same activity is performed in response to the notification of a software and/or hardware upgrade carried out on the subject.

The rules associated with the category of the subject are requested to the corresponding repository at block 412. In response thereto, the engine of the rule repository enters the branch block 414. If no rule for the category of the subject is available the process returns to block 404 directly, in order to repeat the operations described above. Conversely, the process continues to block 416 wherein a copy of these rules is returned to the subject. Referring back to the swim-lane of the subject, the returned rules are logged at block 418.

The process continues to block 420, wherein the rules still to be applied on the subject are extracted from the respective log. The subject then applies each extracted rule directly. At first, the process verifies at block 422 whether

the resource specified in the rule is in an inconsistent condition. If so, the corresponding entry of the state catalog is updated accordingly at block 424; the process then continues to block 426. Conversely, the process descends into block 426 directly. Considering now block 426, the current state of the resource is detected. The process then branches at block 428. If the resource is already in the target state defined in the rule, no action is required and the process returns to block 404. On the contrary, the process continues to block 430 wherein the actions required to bring the resource to the target state from the current state are extracted from the transition table. The actions are executed on the subject at block 432; if the execution of the actions is successful, the rule is flagged accordingly in the respective log. In any case, the process then returns to block 404.

Alternatively, as shown in Figure 4c, the subject operates in a reactive mode. At first, the subject implements a policy for updating the rules applied to the subject, whenever its category changes. The policy consists of a set of activities beginning at the black start circle 442 in the swim-lane of the subject. Proceeding to block 444, the current category (static logic component, dynamic logic component and physic component) of the subject is detected as described above (with reference to blocks 403 and 405-410 of Figure 4a). A test is made at decision block 446 to determine whether the category of the subject has changed since a last verification. If not, the process returns to block 444, in order to repeat the operations described above.

Conversely, the rules associated with the new category of the subject are requested to the rule repository at block 448. In response thereto, the engine of the rule repository logs the request at block 450. Moving to block 452, the activities on the rule repository branch into two mutually exclusive transitions. If no rule for the category of the subject is available, the process returns to block 444 directly. On the

contrary, a copy of these rules is returned to the subject at block 454. The returned rules are logged on the subject at block 456. Proceeding to block 458, each rule is applied by the subject repeating the operations described above (with reference to blocks 420-432 of Figures 4a-4b). The process then returns to block 444.

The actual flow of activities implementing a management process 400c in the reactive mode starts at block 460 in the swim-lane of the authority, wherein one or more new rules are published. In response thereto, the rule repository inserts the new rules into its shared memory space at block 462. A test is made at decision block 464 to determine whether one or more of the new rules match the requests logged in the rule repository. If no matching rule is found, the process returns to block 444. Conversely, a copy of the matching rules is sent to the subject at block 466. The subject logs these rules at block 468; each rule is then applied by the subject at block 469. At the end, the process returns to block 444.

In addition to the pulling mode or to the reactive mode described above (or to both of them), the subject may also implement a healing mode of operation. As shown in Figure 4d, a corresponding method 400d starts at block 472 and then enters a waiting loop at block 474. As soon as a pre-set time-out (for example, 1 hour) has expired, the process descends into block 476 wherein all the rules previously applied to the subject are retrieved from the corresponding log.

Each retrieved rule (starting from the first one) is applied again at block 478. In this way, the resource associated with the rule is first verified, in order to determine whether it is in an inconsistent condition; if so, the corresponding entry of the state catalog is updated accordingly. Therefore, when the current state of the resource is detected no action is carried out if the resource is correctly in the target state specified in the rule. Conversely, whenever the resource features any inconsistency

it is restored to the correct target state by executing the corresponding management actions extracted from the transition table.

Considering now block 480, the method verifies whether the last rule has been processed. If not, the method returns to block 478, in order to apply again a next rule. Conversely, the method returns to the waiting block 474.

For example, let us consider a rule specifying that the resource APPLICATION must be in the state INSTALLED for the category of the subject. When the rule is first applied, the resource APPLICATION is brought to the target state INSTALLED. We now assume that the resource APPLICATION is brought to an inconsistent condition; for example, a component of the application (such as a file) is removed by mistake or during installation of another application. In this case, the verification of the rule detects the inconstancy and sets the state of the resource to INSTALLED_IN_ERROR. As a consequence, the rule is applied again performing the actions required to bring the resource APPLICATION from the state INSTALLED_IN_ERROR to the state INSTALLED.

Moreover, it should be noted that the proposed components defining the category of the subject remove the need to specify any information relating to a possible workflow implying some sort of conditioning or sequencing in the application of the rules; for example, this makes it possible to implement different dependency schemes (such as prerequisites, co-requisites and ex-requisites) without any explicit definition in the rules. For this purpose, it is enough to create rules that specify the target state of each resource for different (physic) categories defined according to the installation of the depending resources.

For example, let us consider a resource APPLICATION to be installed on the subjects of the (static) logic category SECRETARY; the resource APPLICATION requires the installation of a further resource PATCH as a co-requisite. For this purpose, the authority publishes a rule RULE1 for the logic

category SECRETARY specifying that the resource APPLICATION must be in the state INSTALLED, and a rule RULE2 for the physic category APPLICATION=INSTALLED specifying that the resource PATCH must be in the state INSTALLED. A subject of the logic category SECRETARY retrieves and applies the corresponding rule RULE1, causing the resource APPLICATION to be installed. As a result, the same subject moves to the physic category APPLICATION=INSTALLED. Therefore, the rule RULE2 is retrieved and applied, causing the resource PATCH to be installed as well.

As a further example, let us consider a different resource PACKAGE to be installed on the subjects of the same logic category SECRETARY; the resource PACKAGE requires a resource LIBRARY as a pre-requisite. For this purpose, the authority publishes two rules for the logic category SECRETARY; a rule RULE3 specifies that the resource PACKAGE must be in the state INSTALLED, and a rule RULE4 specifies that the resource LIBRARY must be in the state INSTALLED. A subject of the logic category SECRETARY retrieves and applies both the rules RULE3 and RULE4 (in a random order). If the rule RULE3 is applied first, the execution of the required management actions aborts (since the pre-requisite resource LIBRARY is not installed yet) and the resource PACKAGE is not installed (so that the rule RULE3 will be applied again later on); on the other hand, the rule RULE4 is applied correctly. Therefore, the next application of the rule RULE3 is successful so as to bring the subject to the desired configuration.

Similar considerations apply if equivalent methods are envisaged (for example, with error routines that are called on the subject when the number of unsuccessful rules does not decrease at any application loop), if each subject operates both in the pulling mode and in the reactive mode, if the mode of operation of the subject is updated dynamically by corresponding rules, if other policies are envisaged for the

subjects operating in the reactive mode, if each time-out has a different duration, and the like.

Considering now the helper, its operation in the pulling mode is illustrated by the activity diagram of Figures 5a-5b (similar considerations apply to the reactive mode and to the healing mode of operation). A corresponding resource management process 500 begins at the black start circle 503 in the swim-lane of the helper. The helper then enters a waiting loop at block 506. As soon as a pre-set time-out (for example, 1 minute) has expired, the process continues to block 509 wherein the helper inquires an associated dependent object for its category. In response thereto, the subject determines the current category at block 512. Proceeding to block 515, the current category is returned to the helper.

Referring back to the swim-lane of the helper, the rules associated with the category of the subject are requested to the rule repository at block 521. In response thereto, the engine of the rule repository enters the branch block 524. If no rule for the category of the subject is available the process returns to block 506 directly, in order to repeat the operations described above. Conversely, the process continues to block 527 wherein a copy of these rules is returned to the helper. Moving to block 530, the helper logs the returned rules into the corresponding entry of the subject inventory.

The process continues to block 531, wherein the rules still to be applied on the subject are extracted from the log of the subject inventory. Each extracted rule is then enforced on the subject by the helper. Particularly, at block 533 the helper requests the subject to verify the state of the resource associated with the rule. Proceeding to block 536, the subject performs the requested verification; the result of the verification is returned to the helper at block 539. Moving to decision block 542 in the swim-lane of the helper, if the returned result is indicative of any inconsistency the corresponding entry of the state catalog in the subject inventory is updated accordingly at block 545; the process

then continues to block 548. Conversely, the process descends into block 548 directly. Considering now block 548, the current state of the resource is extracted from the state catalog in the subject inventory. The process then branches at block 551. If the resource is already in the target state defined in the rule, no action is required and the process returns to block 506. Conversely, the process continues to block 554 wherein the management actions required to bring the resource from the current state to the target state are extracted from the respective transition table of the subject inventory. Descending into block 557, the helper enforces the management actions on the subject. In response thereto, the subject at block 560 executes the management actions. The process passes to block 562, wherein a code indicative of the result of the management actions is returned to the helper. Considering now block 564 in the swim-lane of the helper, if the management actions have been successfully executed the rule is accordingly flagged in the log of the subject inventory. In any case, the process then returns to block 506.

Similar considerations apply if the helper controls more subjects, if each subject is associated dynamically with different helpers, if the helper and the subject interact in a different manner (for example, with the process that is started by the subject), if a different time-out is used, and the like.

Referring now to Figure 6a, the above-described resource management environment is preferably implemented with a software application 600 that is written in an object-oriented language (such as Java). The application 600 consists of a series of packages, each one grouping logically related classes. A package AUTHORITY 605 depends on a package RULES 610. The package RULES 610 knows a package TUPLE_SPACE 615 containing the specific tuples implementing the rules; the package RULES 610 also depends on a package RESOURCES 620. The package RESOURCES 620 knows a further package MANAGEMENT_ACTIONS 625. At the end, a package SUBJECTS 630

depends on the package RULES 610 (since every subject must know how to retrieve and apply the rules) and on the package RESOURCES 620 (for controlling the corresponding resources under management).

It should be noted that the package RESOURCES 620 and the package MANAGEMENT_ACTIONS 625 are completely unaware of the package RULES 610. In this way, the corresponding contexts are substantially independent; therefore, groups of persons with different roles and skills may be employed in the development of the application 600.

The corresponding class diagram is shown in Figure 6b. A class AUTHORITY 655 represents the authority of the system and a class SUBJECT 658 represents a generic subject. The class AUTHORITY 655 and the class SUBJECT 658 instantiate objects of a class RULE 660, which represents a generic rule of the resource management environment. The class RULE 660 has a private attribute RES_TYPE (specifying the class name of the corresponding resource) and a private attribute TG_STATE (specifying the target state of the resource); moreover, the class RULE 660 exposes a public method APPLY for implementing the corresponding rule on the subject.

The class RULE 660 has a collection consisting of any number of objects instantiating a class PROPERTY 665; each one of these objects includes additional information needed to apply the rule. A class TUPLE_RULE 670 extends the class RULE 660; the class TUPLE_RULE 670 defines a tuple, which represents an actual implementation of the corresponding rule in the tuple space.

The class RULE 660 and the class SUBJECT 658 instantiate objects implementing an interface RESOURCE 675, which represents a generic resource under management. The interface RESOURCE 675 declares an abstract attribute TR_TABLE for the corresponding transition table; moreover, the interface RESOURCE 675 exposes a public method VERIFY (for validating the state of the resource), a public method GET_CURR (returning the current state of the resource), and a public

method CHANGE_STATE (for bringing the resource to the target state from the current state).

The interface RESOURCE 675 is implemented by concrete classes 680 (generically denoted with SPECIFIC_RESOURCE), each one corresponding to a different type of resource (such as a file, a directory, a software package, a monitoring activity, and so on). Each class SPECIFIC_RESOURCE 680 defines the content of the transition table in the attribute TR_TABLE.

The interface RESOURCE 675 is further associated with any number of objects implementing an interface MANAGEMENT_ACTION 685 (which in turn is recursively associated with itself). The interface MANAGEMENT_ACTION 685 exposes a public method EXEC (for performing the required operations on the subject).

The interface MANAGEMENT_ACTION 685 is implemented by concrete classes 690 (generically denoted with SPECIFIC_ACTION), each one corresponding to a different type of management action (such as add, remove or change the attributes of a file/directory, install, remove or restore a software package, trigger or stop a monitoring activity, and so on). Each class SPECIFIC_ACTION 690 actually defines the method EXEC.

More specifically, each tuple is typically represented by a list of fields (up to 16) separated by commas and enclosed in parentheses; each field consists of a key/value pair. The tuple includes a first part (WHO part) that describes the corresponding category; particularly, a key GROUP defines a combination of physic components, static logic components and/or dynamic logic components. A second part of the tuple (WHAT part) describes the corresponding resource. In detail, a key RES_TYPE defines the class name of the resource and a key TG_STATE defines the target state of the resource. The tuple may include further key/value pairs defining additional information necessary to apply the rule; for instance, a key NAME identifies a particular instance of a software component (such as a file name), a key SOURCE identifies an address (such as an URL) from which the software component can be

downloaded, a key DEST identifies the destination (such a folder) of the software component, and so on.

A tuple is added to the tuple space using a primitive OUT. For example, the following tuple:

```
(GROUP:SECRETARY,RES_TYPE:APPLICATION,TG_STATE:INSTALLED)
```

is inserted into the tuple space invoking the primitive:

```
OUT(GROUP:SECRETARY,RES_TYPE:APPLICATION,  
TG_STATE:INSTALLED)
```

The tuples are read using either a non-blocking primitive RD or a blocking primitive IN. In both cases, actual parameters (consisting of values defining a search pattern) and formal parameters followed by a question mark (defining the requested fields) are passed to the tuple space. The formal parameters are used as wildcards to extract the tuples whose content matches the specified pattern (with the question mark alone that denotes all the fields). For example, the following primitive:

```
RD(GROUP:SECRETARY,?)
```

will return the tuple described above.

The tuple space implements a peer-to-peer communication model, which totally decouples the authority from the subjects. Different levels of decoupling are supported. At first, a destination decoupling results from the fact that the authority and the subjects do not need to refer to each other explicitly (thereby providing a fully anonymous communication scheme). In other words, the authority and the subjects do not have any mutual knowledge of their location; conversely, it is the tuple space that identifies the receivers using logic boundaries implied by the information included in any request. Moreover, time decoupling results from the fact that the authority and the subjects do not need to be available at the same time. At the end, space decoupling results from the fact that the tuple is managed in the system with its actual location that is hidden to both the authority and the subjects.

The application of a generic rule involves the instantiation of an object of the type `TUPLE_RULE` (passing the corresponding tuple as a parameter), followed by the calling of the method `APPLY`. The method `APPLY` instantiates an object of the type `SPECIFIC_RESOURCE` (whose name is stored in the attribute `RES_TYPE`). This method in sequence calls the method `VERIFY` and the method `GET_CURR` on the object of the type `SPECIFIC_RESOURCE`; the method `CHANGE_STATE` is then called passing the current state (returned by the method `GET_CURR`) and the target state (stored in the attribute `TG_STATE`) as parameters. The method `CHANGE_STATE` queries the transition table (stored in the attribute `TR_TABLE`), in order to determine the required management actions. One or more objects of the corresponding type `SPECIFIC_ACTION` are instantiated, and the method `EXEC` is called on every one of them.

An example of application of a rule (on a subject operating in the pull mode) is illustrated in the sequence diagram of Figure 7. An object `SUBJECT` starts the sequence of messages requesting the corresponding rules to an object `TUPLE_SPACE`. For this purpose, the object `SUBJECT` calls a method implementing the primitive `RD` (passing its category as a parameter). Assuming that the object `SUBJECT` belongs to the (static logic) category `SECRETARY`, the following primitive is called:

`RD(GROUP:SECRETARY,?)`

The primitive `RD` is not blocking, so that if no rule matching the pattern is available (as in the example shown in the figure) the `SUBJECT` waits for a pre-set period of time before engaging the tuple space again.

An object `AUTHORITY` inserts a new rule into the tuple space calling a method implementing the primitive `OUT`. The rule consists of a tuple specifying that all the subjects of the (static logic) category `SECRETARY` must have a resource `SW_DIST_RES` in a target state `INSTALLED`; the name of a corresponding file (`FILE_NAME`), the address from which the file can be downloaded (`FILE_ADDRESS`) and the folder in which

the file must be installed on the subject (FOLDER_NAME) are specified by corresponding key/value pairs:

```
(GROUP:SECRETARY,RES_TYPE:SW_DIST_RES,  
TG_STATE:INSTALLED,NAME:FILE_NAME,  
SOURCE:FILE_ADDRESS,DEST:FOLDER_NAME)
```

Therefore, when the object SUBJECT queries again the object TUPLE_SPACE a tuple matching the corresponding pattern is available. The object SUBJECT then starts processing the received tuple, and instantiates an object TUPLE_RULE (passing the tuple as a parameter). The sequence of activities continues with the object SUBJECT that calls the method APPLY on the object TUPLE_RULE.

This method instantiates an object SW_DIST_RES (using the class name stored in the attribute RES_TYPE). The method VERIFY and the method GET_CURR are then called in sequence on the object SW_DIST_RES. The process continues with the calling of the method CHANGE_STATE (passing the current state returned by the method GET_CURR and the target state stored in the attribute TG_STATE as parameters). This method determines the required management actions by querying the transition table (stored in the attribute TR_TABLE).

For example, assuming that the file specified in the tuple is not installed on the subject (current state NON_INSTALLED), the transition table indicates that the management actions required to bring the file to the target state INSTALLED from the current state NON_INSTALLED are defined in a class INSTALL. Therefore, an object of the type INSTALL is instantiated; the method EXEC is then called on this object, so as to complete the application of the rule.

Similar considerations apply if the application is implemented in a different language (even non-object-oriented), if the rules have a different structure, if the classes are grouped in different packages, if other classes, interfaces, attributes and/or methods are envisaged, if a different sequence of activities is used to apply the rules, and the like.

The resource management environment described above implements an adaptive model, which can be compared to the behavior of the human society. For example, in the human society a central authority (such as the parliament) enacts a series of laws for the citizens. Each law has a general character, and defines different provisions for corresponding categories of the citizens. For example, financial laws specify a series of tax rates, each one to be applied for a corresponding income range. The parliament publishes the laws using an official bulletin. Every citizen has the responsibility to remain up to date with the laws. In this way, the citizen can calculate the taxes to be paid applying the rate corresponding to his/her income (without the need for the parliament to know the economic situation of every citizen).

Moreover, each citizen may entrust another person with the responsibility of applying the laws (according to his/her personal situation). For example, the citizen may decide to have a certified public accountant take care of all tasks relating to tax payment. In this case, the citizen simply notifies his/her income to the accountant; the accountant calculates the taxes applying the corresponding rate and then informs the citizen of the amount to be paid.

The proposed solution clearly distinguishes from the enforcement model implemented by the resource management environments known in the art. Referring again to the parallel with the civil society, a central authority (such as the gas company) collects information about the conditions of every user; the central authority then issues a series of measures, each one individually tailored to the condition of the corresponding user. For example, the gas company detects the consumption of every user through a meter reading, and then issues corresponding invoices (wherein the amount to be paid is calculated according to the consumption). The invoice is sent to the user, which has a pre-set period of time for paying the amount specified in the invoice.

In other words, the model described above provides an autonomic computing implementation of the resource management environment. The term autonomic comes from an analogy to the autonomic central nervous system in the human body, which adjusts to many situations automatically without any external help. Similarly, in the proposed resource management environment each subject is able to configure, tune and repair itself, as well as anticipate and solve performance problems automatically.

More generally, the present invention is implemented in a data processing structure with a distributed architecture. The structure includes multiple subject entities and one or more authority entities.

The invention proposes a resource management method for self-configuring the subject entities. Each subject entity belongs to one or more categories; moreover, the subject entity controls an instance of one or more resources. The authority entity defines a target state of the resources. The method starts with the step of publishing a plurality of rules under the control of the authority entity; each rule includes an indication of the target state of a resource for a category of the subject entities. Each subject entity retrieves the rules corresponding to its category; each retrieved rule is then applied so as to configure the subject entity according to the target state indicated in the rule.

The solution of the invention provides a resource management method that implements an adaptive model. In this way, the subjects are no longer passive entities, but they actively participate in the corresponding process.

The proposed solution supports cooperation between the authority and the subjects. Therefore, any inconsistencies in the subjects may be avoided.

In the method of the present invention, the authority and the subjects are loosely coupled to each other. In the devised solution, integration of the authority with the subjects is

achieved at the information level (instead of at the procedural level).

Moreover, the devised solution also supports subjects that are not available or off-line.

It should be noted that this result can be achieved reusing components already available in known resource management environments (based on the enforcement model). As a matter of fact, the method of the invention changes the way these components are used, not their operation.

The preferred embodiment of the invention described above offers further advantages.

Advantageously, the rules are published in a shared memory space.

This structure provides a peer-to-peer communication model, which implements a complete decoupling of the authority from the subjects (destination decoupling, time decoupling and space decoupling). The shared memory space is particularly advantageous when the authority and the subjects are spatially dispersed in different locations (even if different applications are not excluded).

In a particular embodiment of the present invention, the subjects operate in a pull mode.

This policy avoids any overload of the shared memory space, since the responsibility of staying abreast of the new rules published by the authority rests entirely with the subjects.

In another embodiment of the invention, the subjects operate in a reactive mode.

In this way, every new rule is immediately provided to the related subjects as soon as it is published (avoiding any delay in its application).

As a further enhancement, the method of the invention also supports a healing mode of operation.

This additional feature allows each subject to detect any inconsistency and to self-repair its configuration.

However, the solution according to the present invention leads itself to be implemented with the subjects that request the rules directly to the authority, with different modes of operation (for example, only the pull one or the reactive one), or even without supporting the healing mode of operation.

Preferably, each rule is applied determining management actions required to bring the resource from the current state to the target state.

The chosen method is very simple, but at the same time effective.

As a further improvement, the resource is verified in order to detect any inconsistency (and its current state is updated accordingly).

The proposed algorithm makes it possible to carry out the operations required by the healing mode in a way that is completely opaque to the whole process.

In any case, the method of the invention is suitable to be implemented applying the rules in another way, without carrying out any verification of the resources, or self-repairing the configuration of each subject in a different manner (for example, with an ad-hoc routine).

Advantageously, the solution according to the present invention is implemented with a computer program application, which is provided on CD-ROM. The application consists of different software modules, which are installed on the computer of the authority and on the computer of every subject. Moreover, it should be noted that each module is suitable to be implemented separately and put on the market even as a stand-alone product.

Alternatively, the application is provided on floppy-disks, is pre-loaded onto the hard-disks, or is stored on any other computer readable medium, is sent to the computers through a network (typically the INTERNET), is broadcast, or more generally is provided in any other form directly loadable into a working memory of the computers.

However, the method according to the present invention leads itself to be carried out with an application having a different architecture, or even with a hardware structure (for example, integrated in a chip of semiconductor material).

Naturally, in order to satisfy local and specific requirements, a person skilled in the art may apply to the solution described above many modifications and alterations all of which, however, are included within the scope of protection of the invention as defined by the following claims.

CLAIMS

1. In a data processing structure with a distributed architecture (100) including a plurality of subject entities (110a) and at least one authority entity (105), a resource management method (300,400a) for self-configuring the subject entities, each subject entity belonging to at least one of a plurality of categories and controlling an instance of at least one resource, the at least one authority entity defining a target state of the resources, wherein the method includes the steps of:

publishing (305-335), under the control of the at least one authority entity, a plurality of rules each one including an indication of the target state of a resource for a category of the subject entities,

each subject entity retrieving (403-418) the rules corresponding to the at least one category of the subject entity, and

applying (420-432) each retrieved rule to configure the subject entity according to the target state indicated in the rule.

2. The method (300,400a) according to claim 1, wherein the step of publishing (305-335) the rules includes inserting (320) the rules into a shared memory space accessible by the subject entities.

3. The method (300,400a) according to claim 2, wherein the step of retrieving (403-418) the rules corresponding to the at least one category of the subject entity includes:

periodically requesting (412) the rules corresponding to the at least one category of the subject entity to an engine controlling the shared memory space, and

returning (416) the rules available in the shared memory space for the at least one category of the subject entity.

4. The method (300,400c) according to claim 2 or 3, wherein the step of retrieving (442-456) the rules corresponding to the at least one category of the subject entity includes:

submitting (448) a request for the rules corresponding to the at least one category of the subject entity to an engine controlling the shared memory space,

storing (450) the request by the engine,

inserting (460-462) a new rule into the shared memory space, and

returning (466-468) the new rule to the subject entity if matching (464) the stored request.

5. The method (300,400a-400d) according to any claim from 1 to 4, further including the steps of:

storing (418;456,468) an indication of the rules applied on the subject entity, and

periodically re-applying (472-480) the applied rules.

6. The method (300,400a-400d) according to any claim from 1 to 5, wherein the step of applying (420-432) each retrieved rule includes:

detecting (426) a current state of the instance of the resource indicated in the retrieved rule,

determining (430) at least one management action required to bring the instance of the resource from the current state to the target state indicated in the retrieved rule, and

executing (432) the at least one management action.

7. The method (300,400a-400d) according to claim 6, wherein the step of applying (420-432) each retrieved rule further includes:

verifying (422) whether the instance of the resource features an inconsistency, and

setting (424) the current state of the instance of the resource to a value indicative of the inconsistency if the result of the verification is positive.

8. A computer program application (203,209,215;227a-236a,251a) directly loadable into a working memory of a data processing structure (100) with a distributed architecture for performing the resource management method (300,400a) of any claim from 1 to 7 when the program is run on the structure.

9. A program product (180) comprising a computer readable medium on which the program application (203,209,215;227a-236a,251a) of claim 8 is stored.

10. In a data processing structure with a distributed architecture (100) including a plurality of subject entities (110a) and at least one authority entity (105), a computer program (203,209,215) directly loadable into a working memory of a computer of an authority entity for performing a resource management method (300,400a) for self-configuring the subject entities when the program is run on the computer, each subject entity belonging to at least one of a plurality of categories and controlling an instance of at least one resource, the at least one authority entity defining a target state of the resources, wherein the method includes the steps of:

publishing (305-335) a plurality of rules each one including an indication of the target state of a resource for a category of the subject entities, for causing each subject entity to retrieve (403-418) the rules corresponding to the at least one category of the subject entity and to apply (420-432) each retrieved rule to configure the subject entity according to the target state indicated in the rule.

11. In a data processing structure with a distributed architecture (100) including a plurality of subject entities and at least one authority entity (105), a computer program (227a-236a,251a) directly loadable into a working memory of a computer of a subject entity (110a) for performing a resource management method (300,400a) for self-configuring the subject entity when the program is run on the computer, each subject entity belonging to at least one of a plurality of categories

and controlling an instance of at least one resource, the at least one authority entity defining a target state of the resources by publishing (305-335) a plurality of rules each one including an indication of the target state of a resource for a category of the subject entities, wherein the method includes the steps of:

retrieving (403-418) the rules corresponding to the at least one category of the subject entity, and

applying (420-432) each retrieved rule to configure the subject entity according to the target state indicated in the rule.

12. In a data processing structure with a distributed architecture including a plurality of subject entities (110a) and at least one authority entity (105), a resource management system (100) for self-configuring the subject entities, each subject entity belonging to at least one of a plurality of categories and controlling an instance of at least one resource, the at least one authority entity including means (203,209,215) for defining a target state of the resources by publishing a plurality of rules each one including an indication of the target state of a resource for a category of the subject entities, and each subject entity including means (227a-236a,251) for retrieving the rules corresponding to the at least one category of the subject entity and means (251a) for applying each retrieved rule to configure the subject entity according to the target state indicated in the rule.

13. In a data processing structure (100) with a distributed architecture including a plurality of subject entities (110a) and at least one authority entity (105), a computer (105) of an authority entity suitable to be used in a resource management method for self-configuring the subject entities, each subject entity belonging to at least one of a plurality of categories and controlling an instance of at least one resource, the at least one authority entity defining a target state of the resources, wherein the computer includes means (203,209,215)

for publishing a plurality of rules each one including an indication of the target state of a resource for a category of the subject entities, for causing each subject entity to retrieve the rules corresponding to the at least one category of the subject entity and to apply each retrieved rule to configure the subject entity according to the target state indicated in the rule.

14. In a data processing structure (100) with a distributed architecture including a plurality of subject entities and at least one authority entity (105), a computer (110a) of a subject entity suitable to be used in a resource management method for self-configuring the subject entities, each subject entity belonging to at least one of a plurality of categories and controlling an instance of at least one resource, the at least one authority entity defining a target state of the resources by publishing a plurality of rules each one including an indication of the target state of a resource for a category of the subject entities, wherein the computer includes means (227a-236a,251a) for retrieving the rules corresponding to the at least one category of the subject entity, and means (251a) for applying each retrieved rule to configure the subject entity according to the target state indicated in the rule.

AN ADAPTIVE RESOURCE MANAGEMENT METHOD AND SYSTEM

Abstract

A resource management method (400c) and a corresponding system implementing an adaptive model are proposed. The method allows an authority to define a desired target state of several kinds of resources, which are directly controlled by different subjects. The authority publishes (460) a series of rules in a shared repository; each rule indicates the target state of a resource for a corresponding category of the subjects. Each subject retrieves (444-456;462-468) the rules corresponding to its category from the shared repository. The rules are then applied (458;469) by the subject directly. In this way, the subjects self-adapt to the rules, without requiring any direct interaction with the authority.

FIGURE 4C

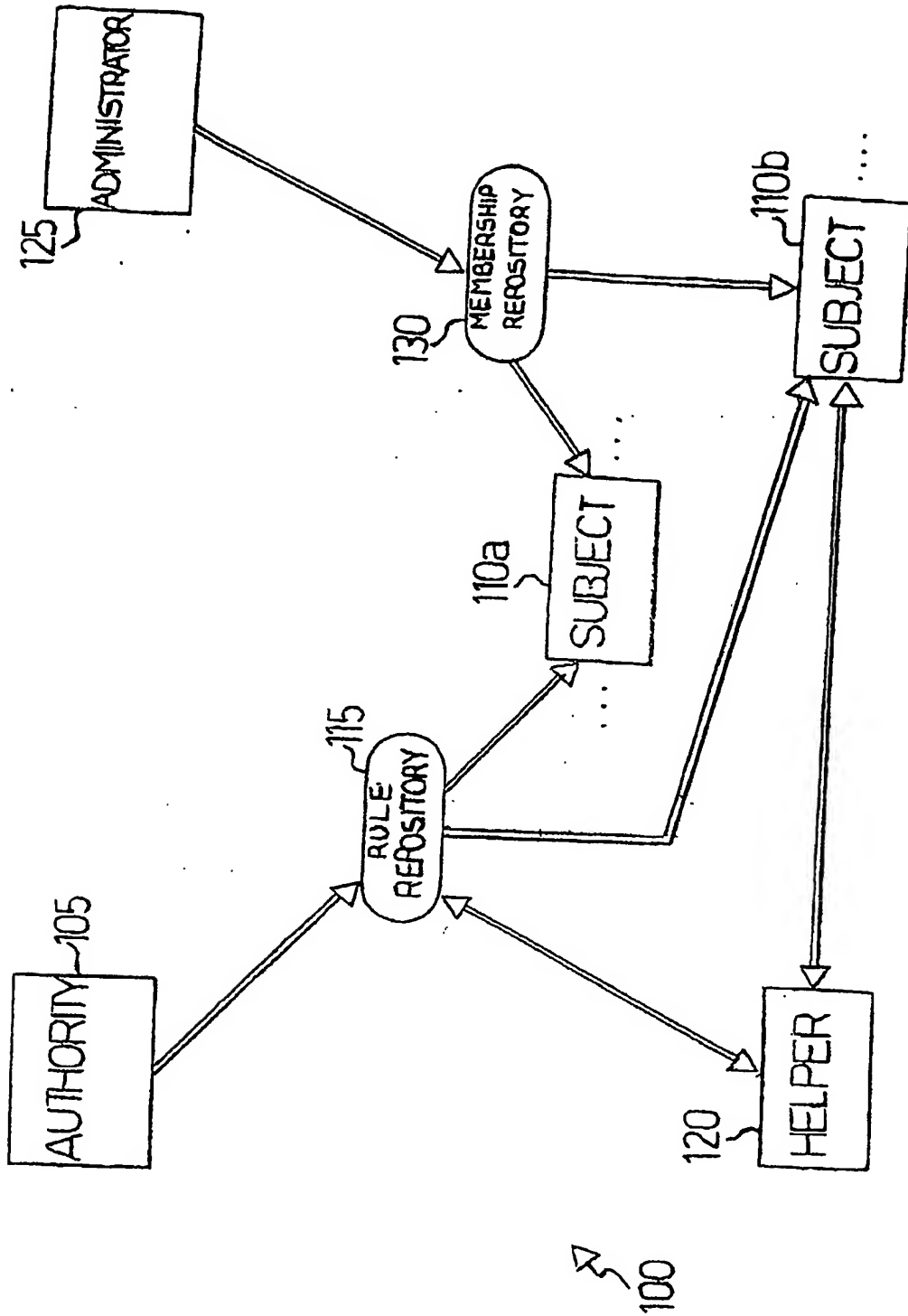


Fig. 1a

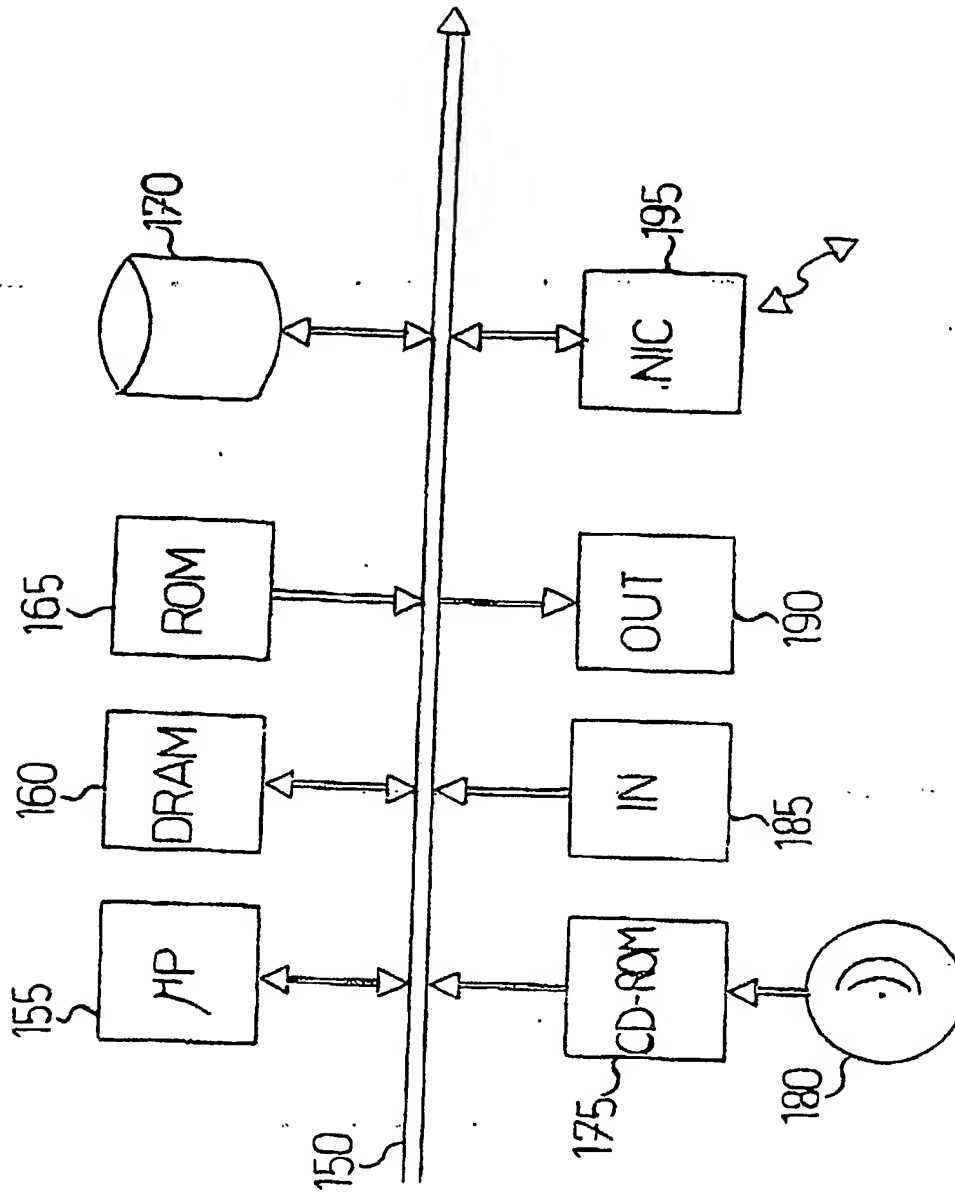


Fig 1b

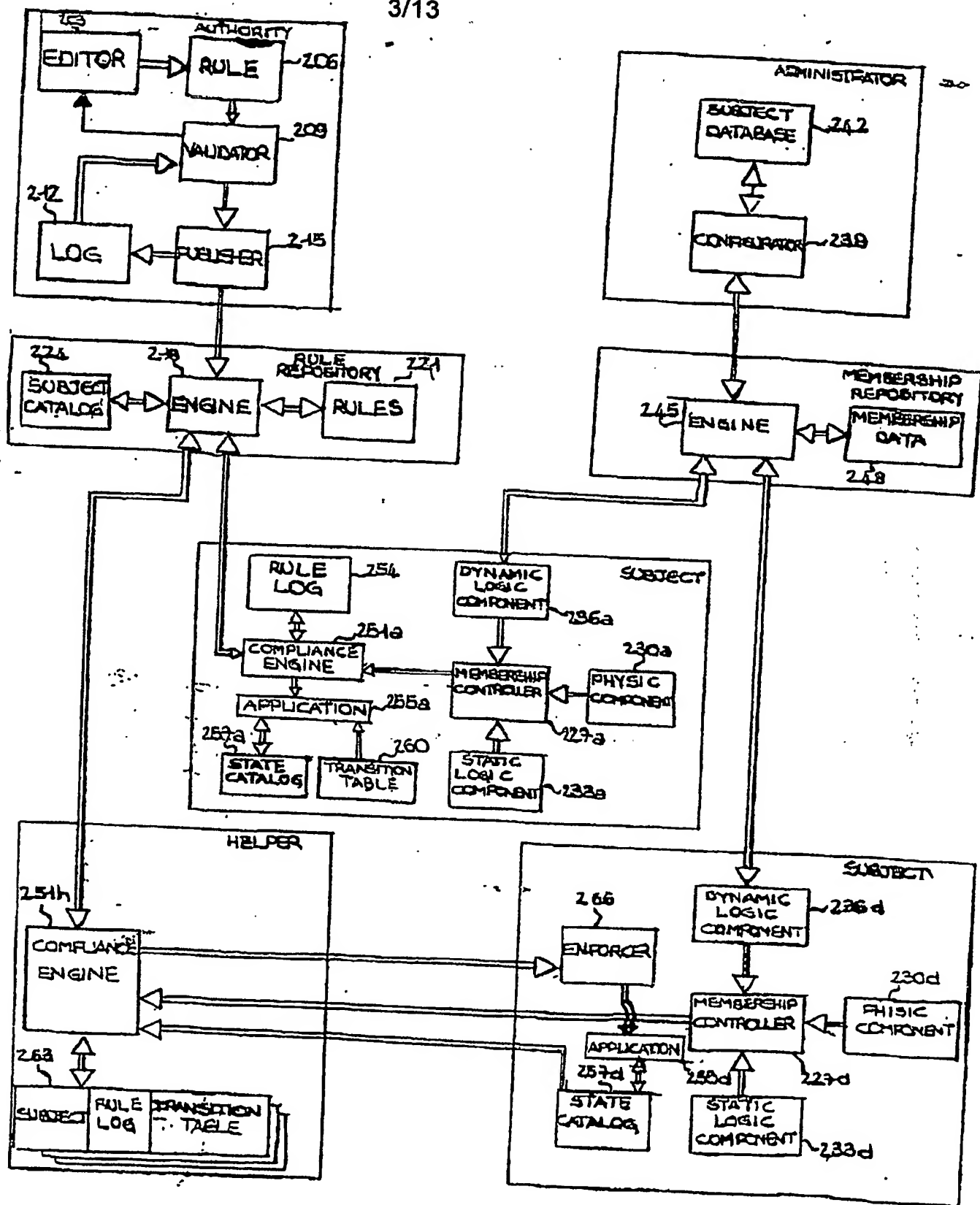


Fig. 2

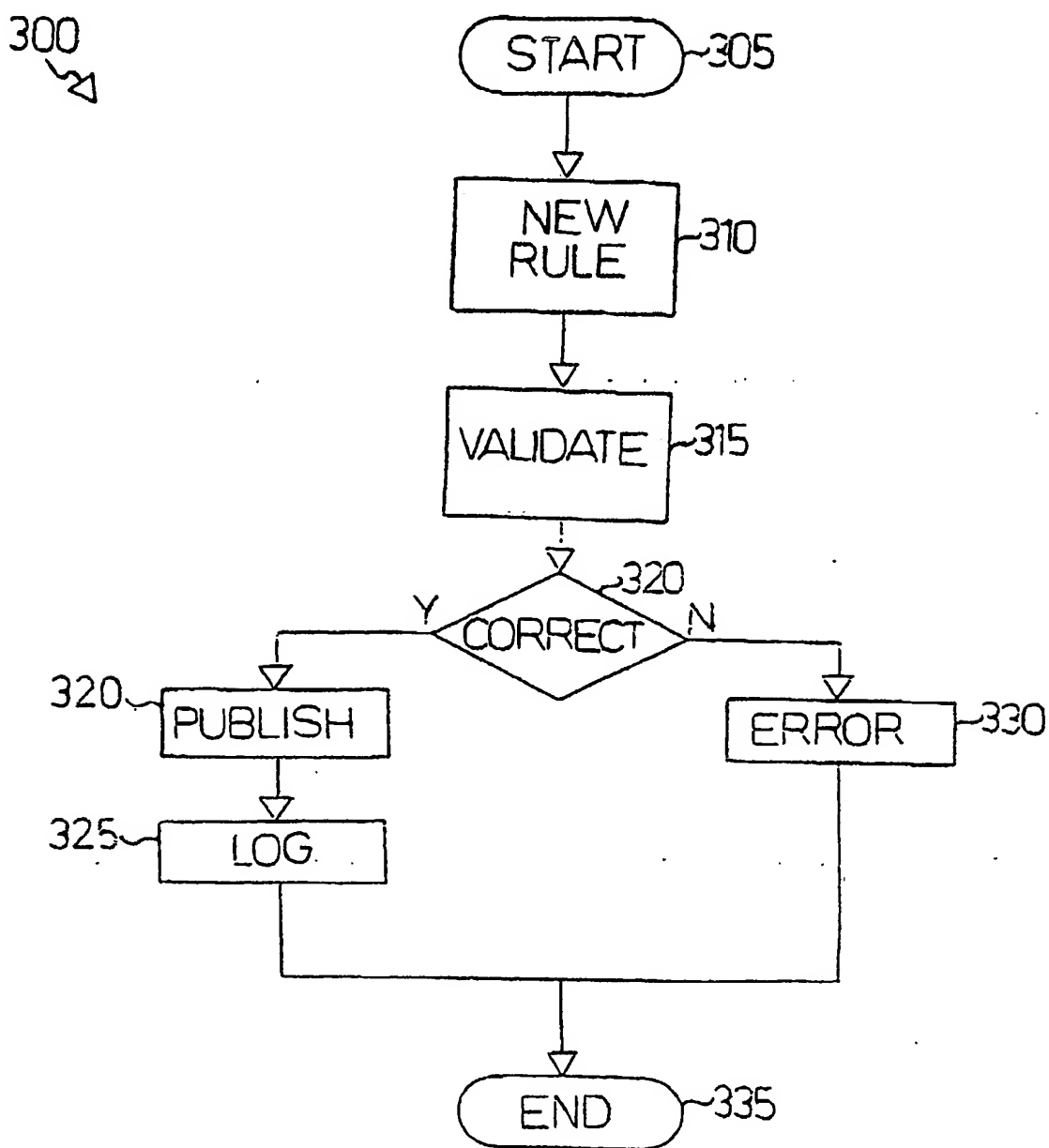


Fig 3

4002

5/13

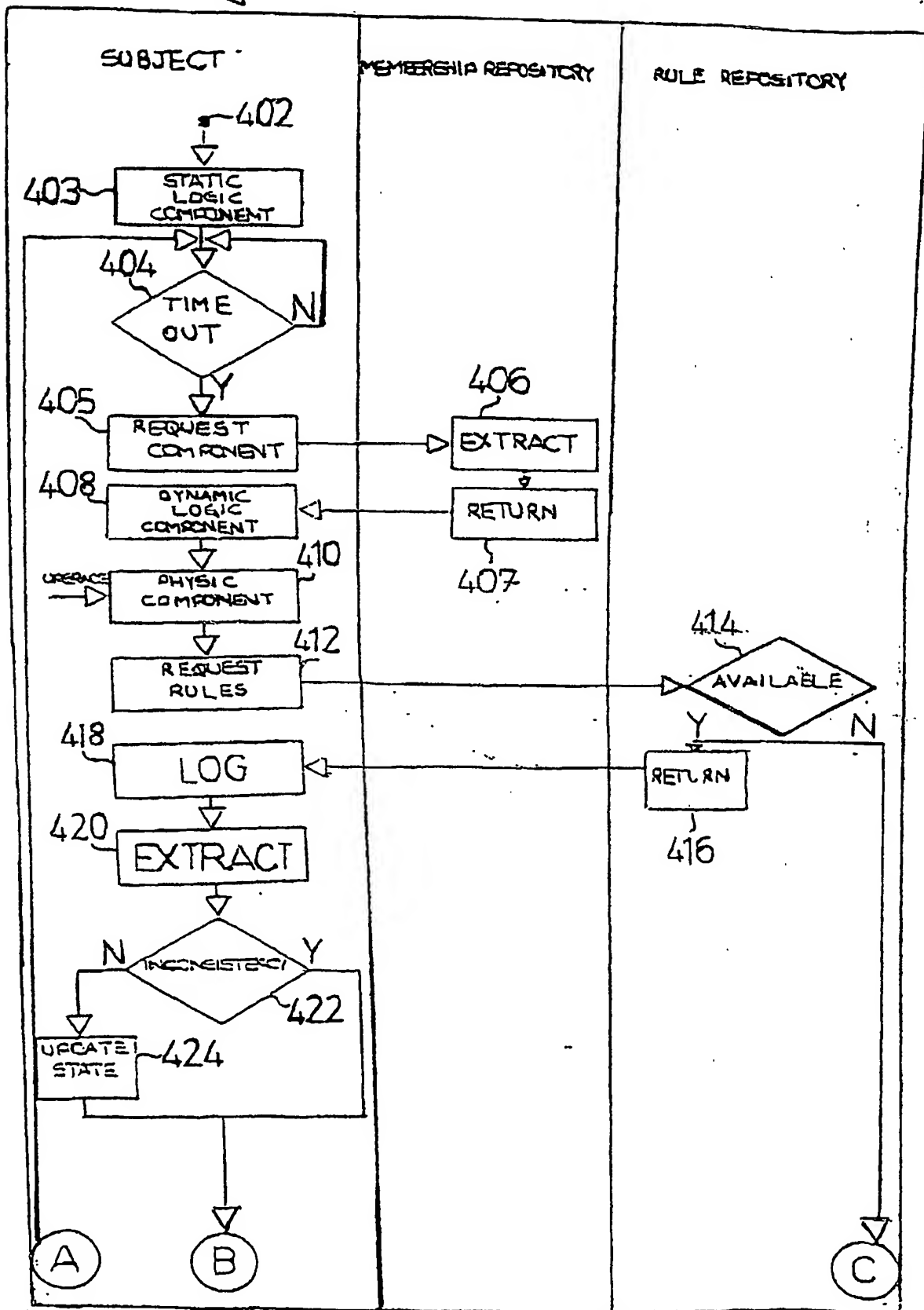
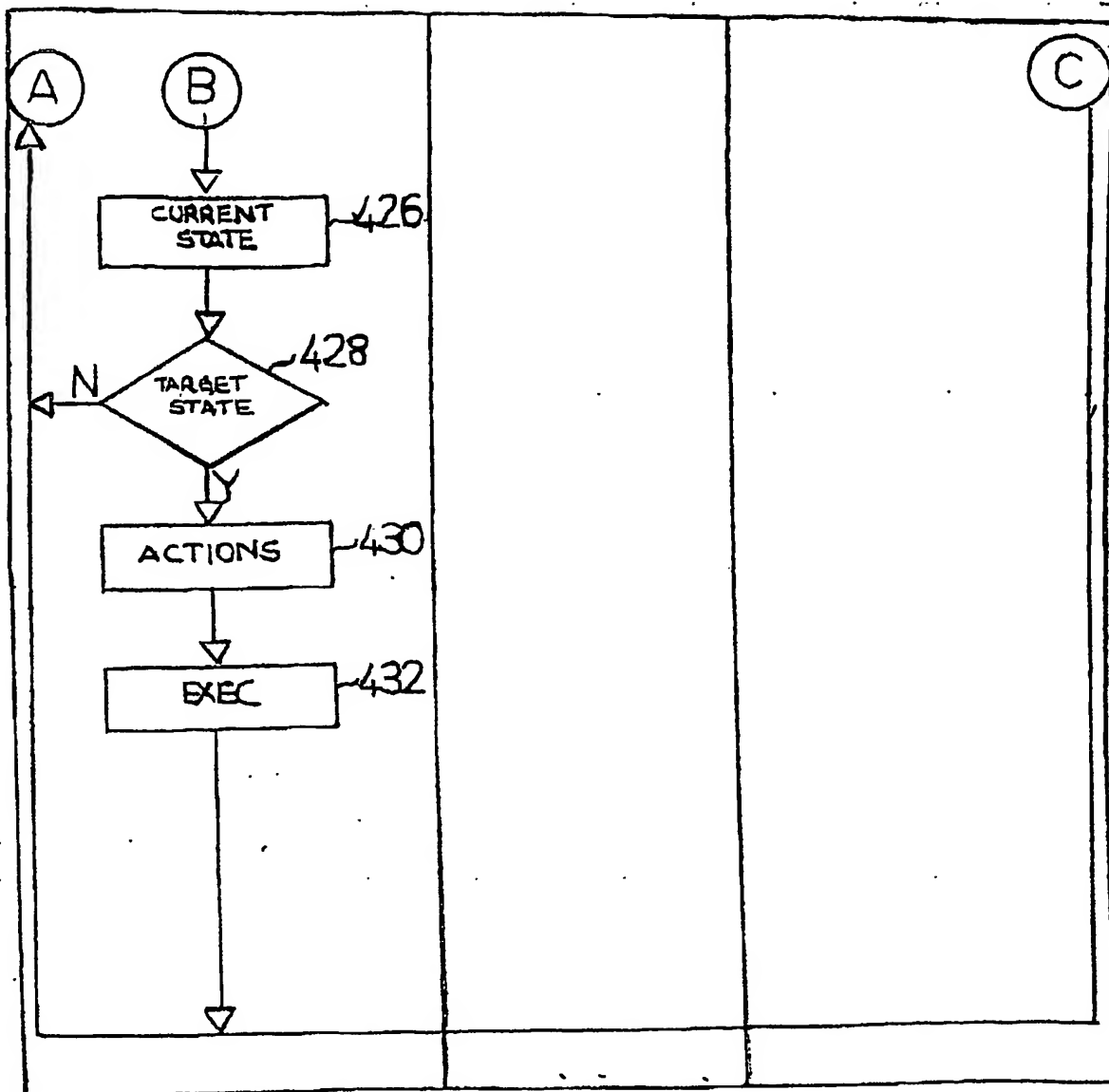


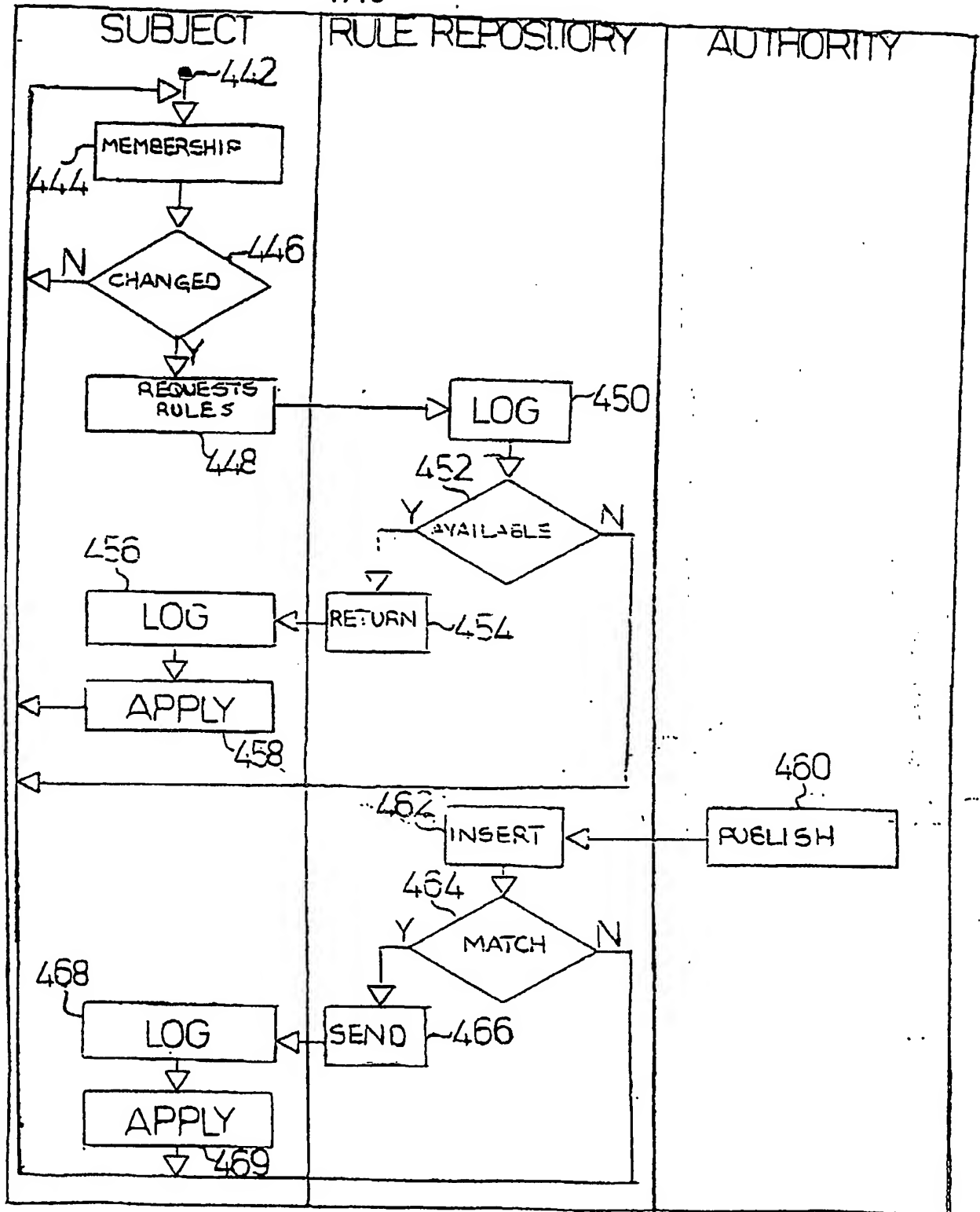
Fig. 4a



400b

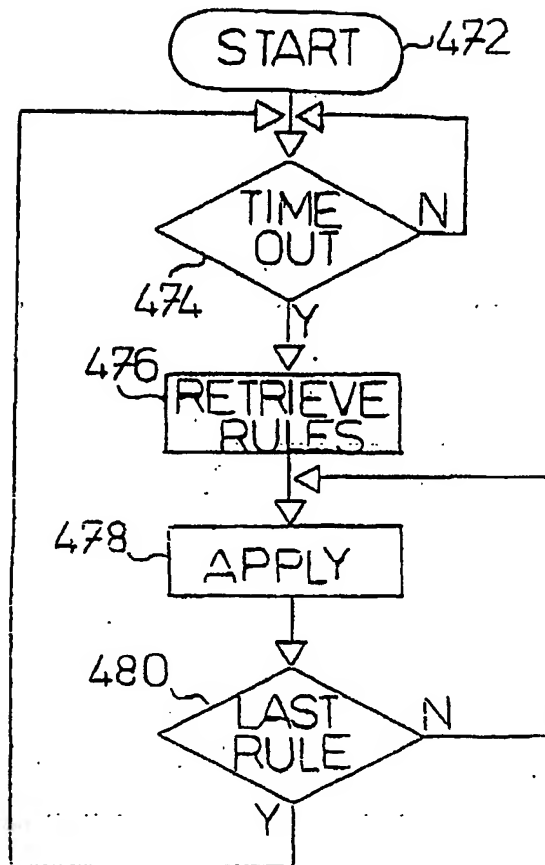
Fig 4b

7/13



400c

Fig. 4c



400d

Fig. 4d

500

9/13

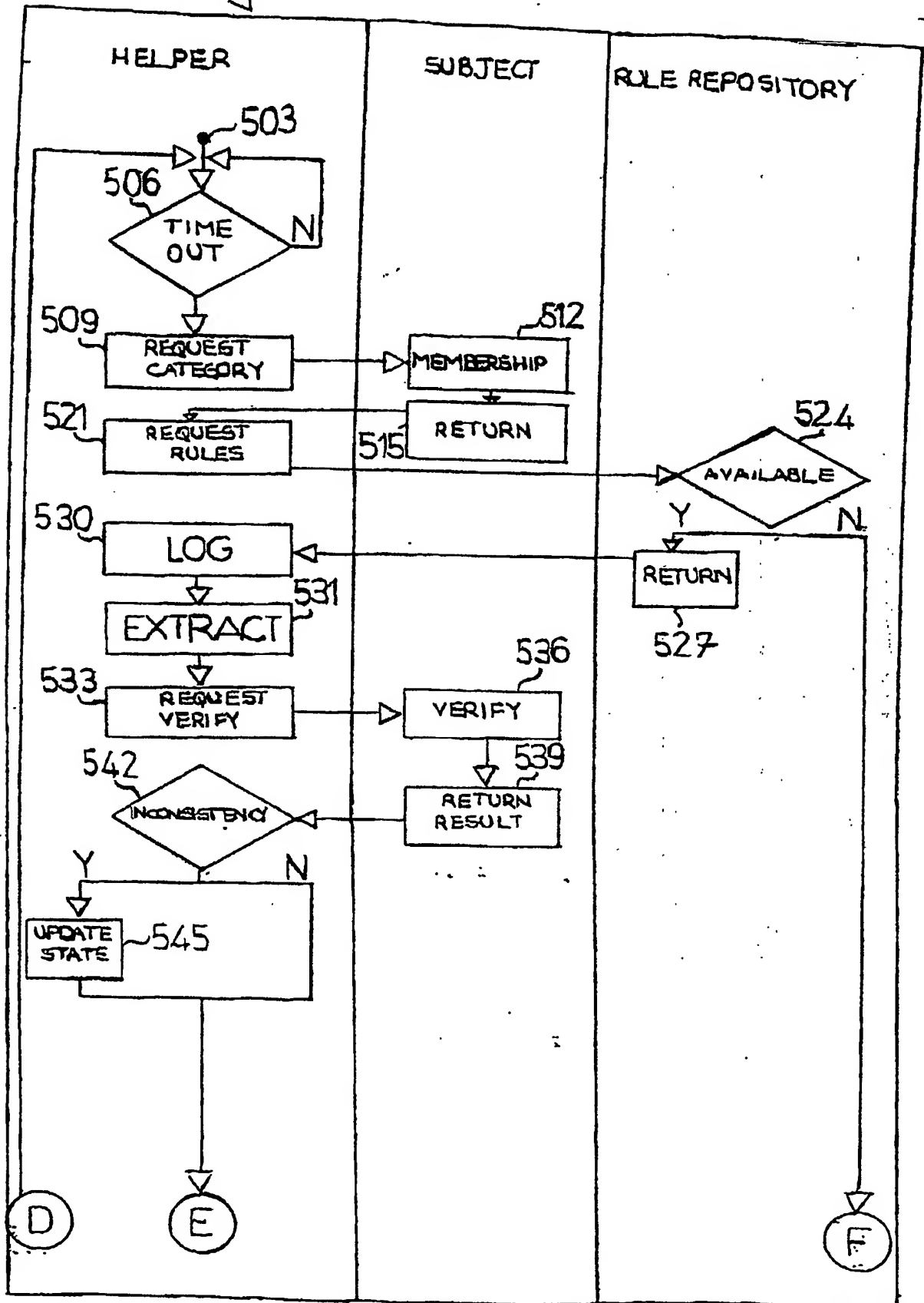
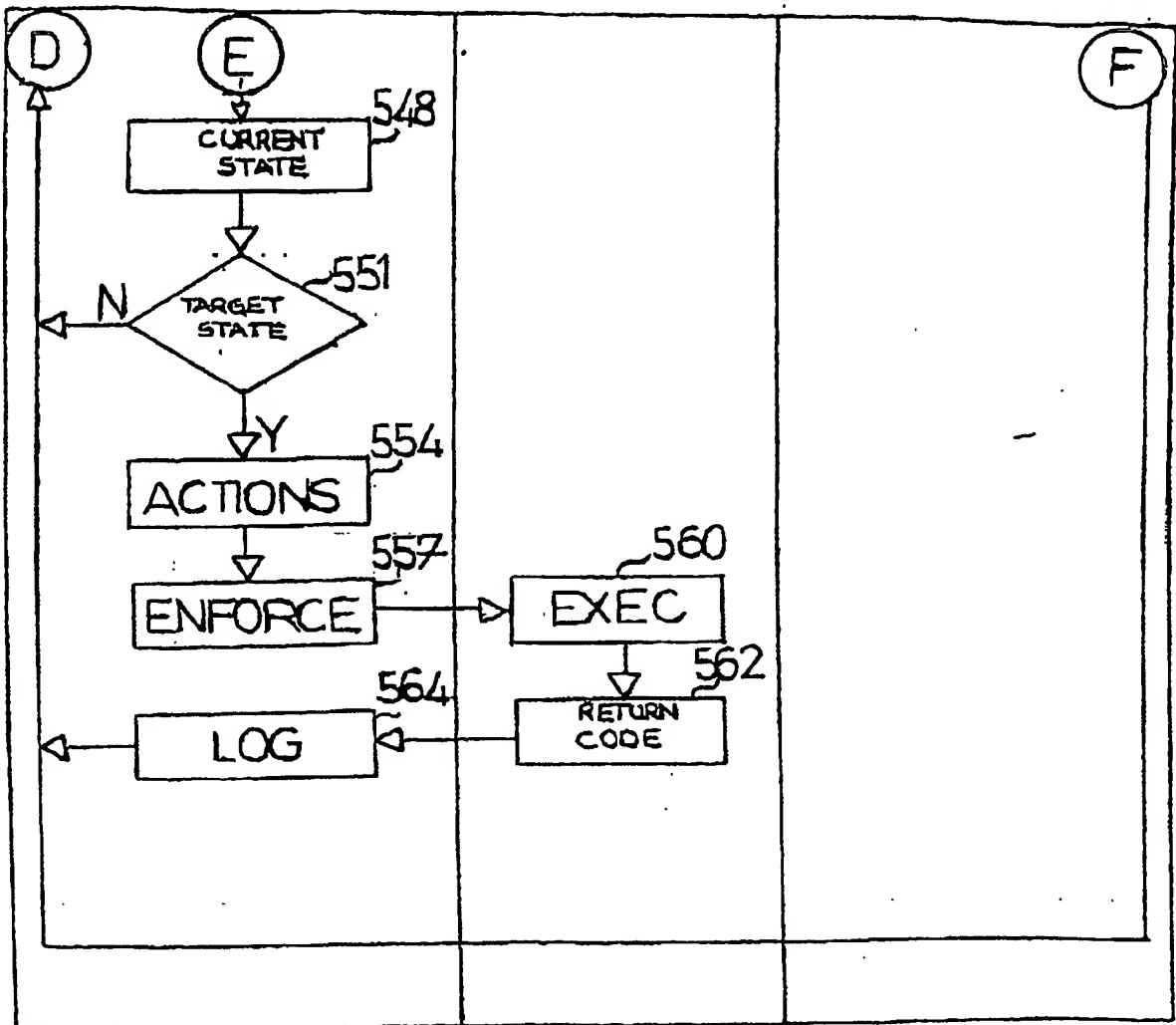


Fig. 5a

10/13



500

Fig 5b

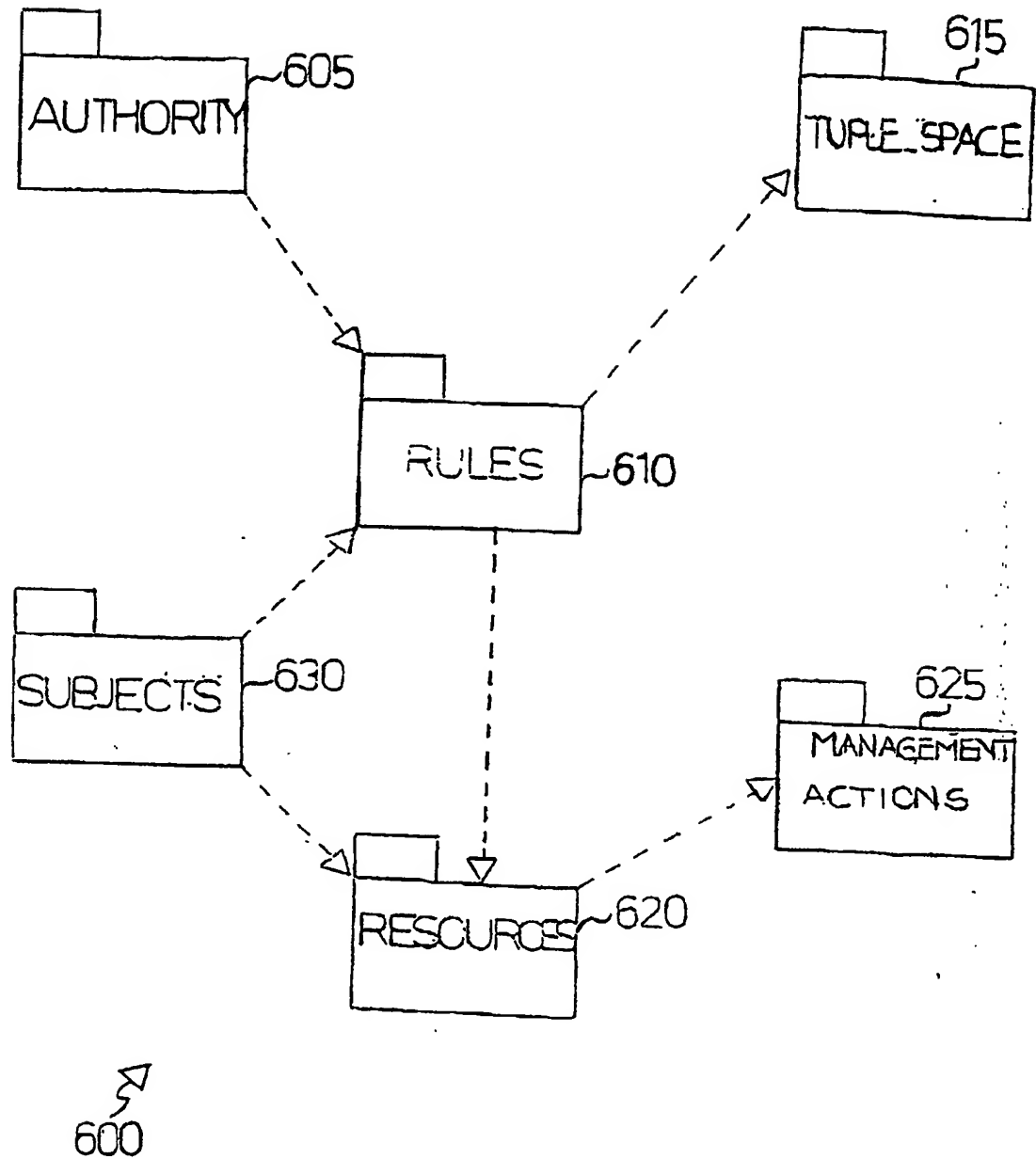


Fig 6a

12/13

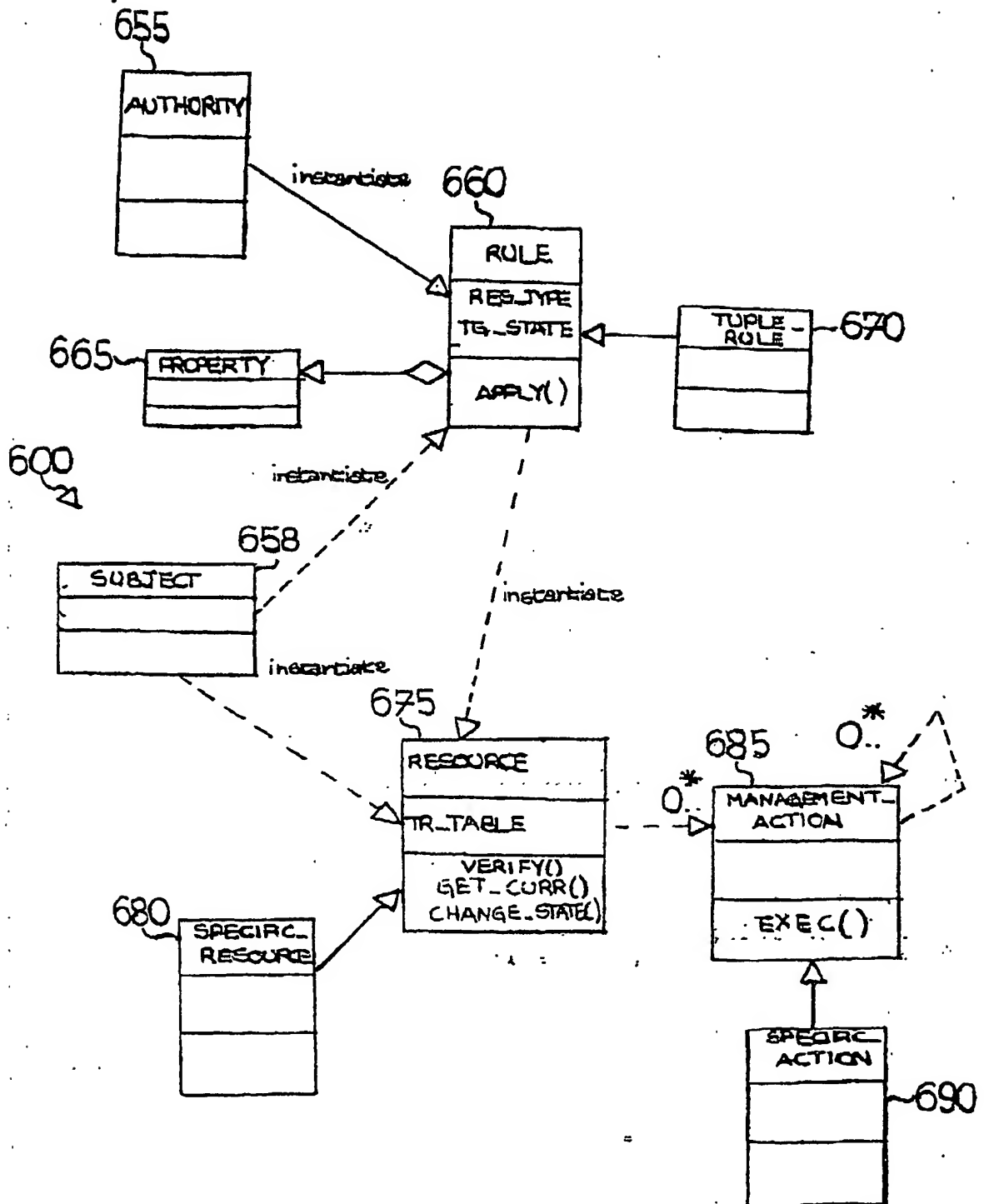


Fig 6b

13/13

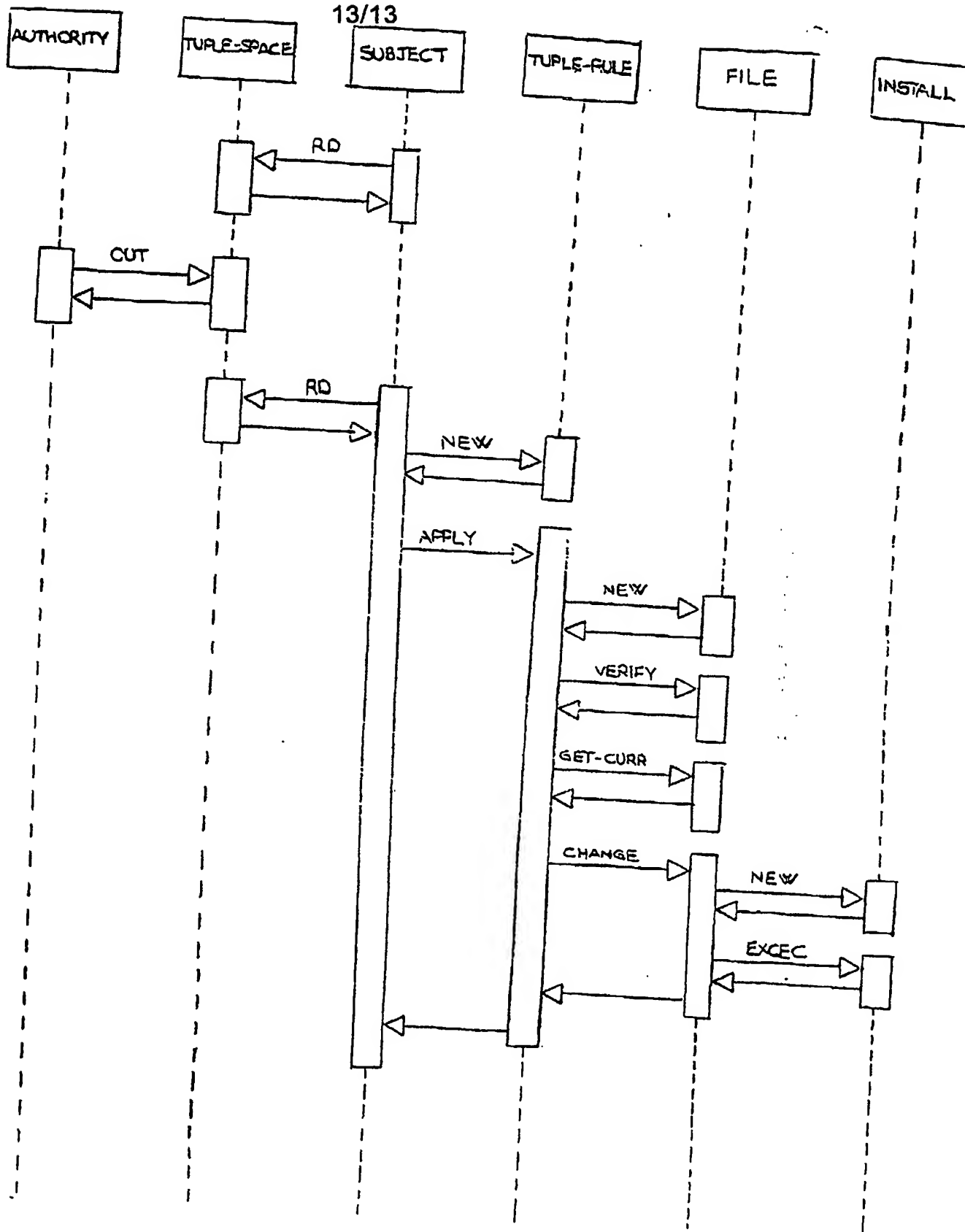


Fig 7

PCT Application
EP0310081



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.